Petit guide utile à l'élaboration d'examens à choix multiples dans le domaine médical

de Foy Th[1], Dory V[1] & Degryse J[2]

Correspondance:
Dr. Thérèse de Foy
UCL
Tour Pasteur, UCL 5360,
Avenue E.Mounier, 53
1200-Bruxelles
Tél: 32-2-7645344

[1] Médecins et chercheurs au CAMG de l'UCL

[2] Professeur en Médecine UCL/ KUL - Réseau de Pédagogie Médicale

Table des matières

1 INTRODUCTION 4

2 DIFFÉRENTS TYPES DE QUESTIONS 5

3 L'ÉVALUATION : UN ÉLÉMENT IMPORTANT DE LA FORMATION 8

4 ELABORATION DES QUESTIONS À CHOIX MULTIPLES 10

4.1 Exigences générales 10

4.2 Différences entre les questions de type A+ et celles de type R 10

4.3 Recommandations pour la rédaction des QCM type A et type R 13

4.3.1 Le thème 13
4.3.2 Le (les) item(s) 13
4.3.2.1 Structure de l'item 13
4.3.2.2 Différents thèmes possibles pour la rédaction d'items en rapport avec les tâches du médecin (référence n°1) 13
4.3.2.3 Règles à respecter lors de la rédaction des items 14
4.3.2.4 Règles supplémentaires à respecter lors de la rédaction des items pour des QCM à appariement étendu 15
4.3.3 L'instruction 15
4.3.3.1 Règles à respecter lors de la rédaction d'une instruction. 15
4.3.3.2 Exemples d'instructions bien formulées. 15
4.3.4 La liste de propositions 16
4.3.4.1 Règles à respecter lors de la rédaction de la liste de propositions 16
4.3.4.2 Différence entre QCM de type A et QCM de type R 16
4.3.4.3 Exemples de données pour la rédaction d'une liste de propositions 16
4.3.5 Défauts techniques à éviter lors de la rédaction de QCM (manuel de Case et Swanson, p 26, référence n°1) 17
4.3.5.1 Les défauts qui fournissent des indices (cues) pour identifier la bonne réponse (ou éliminer les distracteurs) et leurs « remèdes » 17
4.3.5.2 Les défauts liés à des difficultés inutiles inclues dans les questions 18
4.3.5.3 Exemples de QCM mal rédigés 19

5 COMMENT RÉDIGER UN EXAMEN DE TYPE QCM ÉTAPE PAR ÉTAPE ? . 22

5.1 Définitions préalables 22

5.2 Conception d'un plan directeur ou tableau de spécification 22

5.3 Etablissement concret du contenu du plan directeur 22

5.4 Formation d' un groupe d'enseignants à la rédaction d'items 23

5.5 Rédaction des items, vérification, documentation 23

5.6 Vérification en groupe des items 23

5.7 Vérification par des tiers et des experts 23

5.8 Acceptation ou refus des items 24

5.9 Elaboration de l'examen 24

6 INTERPRÉTATION DES RÉSULTATS OBTENUS EN VUE DE L'ANALYSE DES ITEMS 25

7 COMMENT ÉTABLIR UNE NOTE DE PASSAGE ? 28

7.1 Définition et considérations générales 28

7.2 Méthode relative 28

7.3 Méthode absolue 28

7.4 Compromis entre une méthode absolue et une méthode relative 29

8 CONCLUSION 30

9 BIBLIOGRAPHIE 30

10 ANNEXES 31

10.1 Formulaire pour la rédaction d'items (manuel de Case et Swanson, p 93, référence n°1) 31

10.2 Aide-mémoire utile 32

10.3 Exemple d'instruction standardisée pour un QCM de type R 33

10.4 Exemples de QCM bien rédigés 34

10.4.1 QCM à appariement étendu : 34

10.4.2 QCM de type A+ : 36

1 Introduction

Ce petit guide vous aidera lors de l'élaboration de questions à choix multiples. Il est essentiellement basé sur le manuel écrit par Case et Swanson (référence n°1 dans la bibliographie) ainsi que sur le mode d'emploi écrit par Krebs (référence n°2 dans la bibliographie). Ces deux documents sont disponibles librement sur internet (cfr bibliographie pour les liens).
Ce manuel est volontairement dirigé sur l'élaboration des QCM du type « meilleure réponse possible ». En effet, ce type de questions a largement fait ses preuves de validité et fiabilité (si, toutefois, les règles de rédaction sont respectées) et devrait être favorisé lors de toute évaluation écrite.
Après avoir brièvement rappelé l'importance de toute évaluation dans le cursus des études de médecine, nous détaillons les règles de rédaction à respecter lors de l'écriture de QCM.
Ensuite, nous reprendrons les étapes nécessaires à la réalisation d'un examen par QCM.
L'interprétation des résultats obtenus par les étudiants est très importante pour la vérification et le contrôle des questions. Ce point sera abordé au chapitre 6.
Nous terminerons par l'explication de la détermination des cotes ne dit on pas notes de passage avec les différentes méthodes existantes : absolue, relative ou celle basée sur un compromis.

2 Différents types de questions

Il existe deux grands types de questions pour l'évaluation écrite : le type « meilleure réponse possible » et le type « vrai/faux ». Dans chacun de ces types de questions, il existe des variantes.

Le type de questions « vrai/faux » exige de l'étudiant une tâche supplémentaire. Il doit décider jusqu'à quel point une réponse doit être vraie pour être considérée comme telle. Or, cette considération additionnelle n'est probablement pas en relation avec l'expérience et la connaissance cliniques. Il s'agit souvent de deviner ce que le rédacteur pensait car les propositions ne sont pas totalement fausses ou totalement vraies. Case et Swanson, dans leur manuel (référence n°1), déconseillent l'utilisation des questions de type « vrai/faux ».

Le type de questions « meilleure réponse possible » est, par contre, très fiable. Le nombre de réponses à donner doit être indiqué. En général, il n'y a qu'une seule réponse à choisir. Les différents types qui existent sont les suivants (manuel de Krebs, p 5, référence n°2) :

- A+ : choix simple positif parmi cinq propositions de réponse

Exemple :
Un homme de 64 ans se trouve à l'hôpital avec un emphysème pulmonaire au stade terminal. Ces proches constatent qu'il les reconnaît, qu'il est lucide et bien orienté, mais qu'il ne s'intéresse plus autant à leurs visites et qu'il semble très passif.
Quelle est l'explication la plus probable du comportement de ce patient ?
A. évolution vers un délire
B. évolution vers un trouble psychotique
C. excès de sédation
D. péjoration de trouble de la personnalité
E. repli sur soi

- A- : choix simple négatif parmi cinq propositions de réponses

Exemple :
*Lequel des antibiotiques suivants **ne doit pas** être administré à un enfant âgé de 18 mois souffrant d'une otite moyenne aiguë ?*
A. amoxicilline
B. céfaclor
C. co-trimoxazole
D. doxycycline
E. érythromycine

- B : attribution de la réponse correcte parmi cinq propositions de réponse possibles à plusieurs questions

Exemple :
A. le captopril
B. le diazoxide
C. le dinitrate d'isosorbide
D. la nifédipine
E. le propanolol
Quelle substance inhibe
1. la libération de la rénine ?
(A) (B) (C) (D) (E)
2. la dégradation de la bradykinine ?
(A) (B) (C) (D) (E)
3. l'ouverture des canaux calciques ?
(A) (B) (C) (D) (E)

- R : comme le type B, mais avec jusqu'à 26 propositions de réponse (A-Z)

Exemple :
***Thème** : effets secondaires des antihypertenseurs*
***Options** :*
A. Captopril (Capoten®)
B. Celiprolol (Selectol®)
C. Dihydropyridines (Adalat®, Amlor®…)
D. Indapamide (Fludex®)
E. Losartan (Cozaar®, Loortan®)
F. Propanolol (Indéral®)
G. Spironolactone (Aldactone®, Canrenol®…)
H. Vérapamil (Isoptine®, Lodixal®…
***Instruction** :*
Choisissez pour chaque patient ci-dessous le traitement le plus vraisemblablement responsable de la symptomatologie.
***Vignettes cliniques** :*
1. Charles, un homme de 76 ans, souffre depuis 10 ans d'hypertension artérielle. Il se plaint d'une toux sèche que rien n'arrête et qui survient surtout la nuit. L'examen clinique est négatif.
2. Un confrère a prescrit hier à Florentine, 54 ans, un traitement antihypertenseur. Vous êtes appelé à son chevet pour une crise d'asthme d'apparition soudaine.
3. Gustave, 65 ans, est traité pour de l'hypertension. Hier, on lui a fait une prise de sang. On a remarqué une hypokaliémie importante (potassium à 2,1).
4. Mia, 67 ans, est traitée depuis 2 semaines pour hypertension. Elle vient consulter pour des palpitations cardiaques et un gonflement important des chevilles.

- PickN : comme le type A+ ou le type R, mais avec un choix portant sur plus d'une « meilleure réponse possible »

Exemple :

A. acide folique
B. calcium
C. fer
D. fluorure
E. Vitamine A
F. Vitamine B1 (thiamine)
G. Vitamine B6
H. Vitamine B12 (cyanocobalamine)
I. Vitamine C
J. Vitamine D
K. Vitamine E

Choisissez pour chaque enfant les suppléments vitaminés et électrolytiques appropriés.
1. Un nourrisson de 1 mois est amené en contrôle chez le médecin. Il est exclusivement nourri par allaitement et l'examen est parfaitement normal lors de la consultation. (Choisissez 2 compléments alimentaires.)
2. Une fillette de 6 ans souffre de mucoviscidose. Elle ne prend aucun médicament. (Choisissez 3 compléments alimentaires.)°

Les types A+ et R sont ceux recommandés actuellement pour la rédaction d'examen dans le domaine médical. Ce sont ces deux types que nous développerons dans ce petit guide.

3 L'évaluation : un élément important de la formation

L'évaluation a une importance primordiale dans le cursus de l'étudiant. Elle est une grande source de motivation pour les étudiants. L'examen doit rencontrer les buts éducationnels (soit à l'ISEI les PC) fixés par l'enseignant. Il doit tester les connaissances spécifiques jugées primordiales mais aussi la capacité de les mettre en pratique face à une situation concrète. L'importance des caractéristiques psychométriques dépend de l'objectif de l'évaluation. Elles sont primordiales si l'examen a des conséquences majeures. Il s'agit alors de faire un examen de bonne qualité.

Les quelques notions théoriques reprises ci-dessous ont été reprises de la thèse de Degryse J. (référence n°5).

La qualité d'un examen (Q) dépend de sa fiabilité (F), de sa validité (V), de son impact pédagogique (I), de faisabilité (temps requis/le coût) (f) et de son acceptabilité (A).
L'équation suivante résume cela.
Q = F x V x I x f x A

La fiabilité (la mesure faite reflète bien la réalité, elle est exacte, reproductible, les sources d'erreur sont limitées au maximum) est une condition nécessaire mais insuffisante à la validité (le test utilisé mesure bien ce qu'il est censé devoir mesurer).
Il faut distinguer 5 types de validité.
La validité « apparente » (face validity) n'est qu'apparente. Un examen qui a une validité apparente donne l'impression à un observateur de mesurer ce qu'il est censé mesurer.
La validité de contenu (content validity) vérifie que le contenu de l'examen est bien en accord avec ses objectifs et qu'il est représentatif des domaines de compétences qu'il est censé mesurer.
La validité structurelle (construct validity) d'un test reflète la cohérence des résultats avec les hypothèses à la base de sa construction. On s'attend par exemple à ce que des experts du domaine testé aient des meilleurs résultats que des novices.
La validité concomitante (concurrent validity) est évaluée par la corrélation entre le test étudié et un autre test considéré comme une référence valide pour le même domaine.
La validité prédictive (predictive validity) d'un test est sa capacité à prédire une performance à venir. Par exemple, le score à un examen de fin de formation permet-il de prédire quels seront les meilleurs professionnels (pour un critère donné en lien avec le domaine testé) ?

L'évaluation avec des QCM de type A et R permet une bonne fiabilité et une bonne validité si ceux-ci sont rédigés de manière correcte et contrôlés.

Pour une bonne rédaction d'un examen, il faut commencer par écrire un « blueprint » (= « plan directeur » selon la traduction utilisée par Le Collège royal des médecins et chirurgiens du Canada (référence n°4) ou « tableau de spécification »). Avant tout, il faut avoir une définition claire des objectifs d'apprentissage sous forme d'un catalogue par exemple. Avec celui-ci, il y a lieu d'établir une liste pondérée des matières. Ceci permettra que les mesures faites à un examen soient comparables d'une session à l'autre. Les contenus testés seront équivalents et la répartition des thèmes aussi.

LA VALIDITE

(qualité d'un instrument à mesurer réellement ce qu'il doit mesurer en fonction de l'utilisation que l'on veut en faire)

Apparence	Apparente	*Face validity*	Aspect visible extérieur
Interne	Contenu	*Content validity*	Congruence avec objectifs Représentativité des items
	Structurelle	*Construct validity*	Cohérence avec schéma théorique
Externe	Concomitante	*Concurrent validity*	Corrélation avec mesures similaires
	Prédictive	*Predictive validity*	Corrélation avec variables à prédire

4 Elaboration des questions à choix multiples

4.1 Exigences générales

La validité et la fiabilité d'un examen dépendront de la qualité de la rédaction de ses questions c'est-à-dire des items dans le cas de QCM de type « meilleure réponse possible ».
Différentes règles de base sont à respecter.

- Le thème de la question correspond aux exigences futures de la formation et/ou de la pratique professionnelle. Les items devraient donc reprendre les aspects d'un thème important auquel l'étudiant sera souvent confronté dans sa pratique courante et/ou qui pourrait avoir des conséquences graves en cas d'erreur et/ou où les fausses idées sont répandues et/ou qui est essentiel pour la compréhension de futurs domaines d'enseignement. (manuel de Krebs, p 4, référence n°2) Il est important de garder à l'esprit l'importance dans la formation médicale de garder une vision globale avant de se spécialiser.
- Il est demandé à l'étudiant d'appliquer des connaissances dans des situations pratiques. Il ne s'agit pas seulement d'examiner le niveau de connaissances mais aussi d'évaluer la compréhension (analyser, mettre en relation, intégrer, transformer des informations, tirer des conclusions) ainsi que la capacité de mise en pratique du savoir acquis et la capacité d'évaluer des informations. L'utilisation de vignettes cliniques (patient décrit avec ses symptômes, le résultat de l'examen clinique et, éventuellement, d'examens complémentaires) rencontre cette exigence.
- La difficulté de l'item est adaptée au niveau du groupe cible.[3]
- L'item se focalise sur un contenu ou un problème bien défini dans un domaine bien circonscrit. La question est claire.
- Il existe manifestement une solution bien meilleure que les autres. Le bon étudiant peut répondre à la question sans regarder les différentes propositions de réponse.
- Les distracteurs (c'est-à-dire les propositions données qui ne sont pas la réponse correcte) doivent être homogènes (tous sont soit des diagnostics, soit des traitements, soit des pronostics…)
- Eviter les défauts techniques qui permettent aux étudiants habiles de répondre aux QCM mal rédigés sans connaître la réponse. (cfr paragraphe 4.3.5.1.)
- Le questionnaire doit être vérifié par des confrères. S'ils manifestent des difficultés à déterminer la réponse correcte, il faut modifier la liste de propositions ou l'item pour éliminer l'ambiguïté.

4.2 Différences entre les questions de type A+ et celles de type R

La différence entre les questions de type A+ (choix simple positif parmi cinq propositions de réponse) et les questions de type R (attribution de la réponse correcte parmi 6 à 26 propositions (A-Z) de réponse possibles à plusieurs questions) est essentiellement structurelle.

[3] Sauf s'il s'agit de la conception d'un « test de progression ». Dans ce cas particulier, le même examen sera donné à tous les étudiants, tous les niveaux confondus.

Dans le **type A+**, l'énoncé est composé dans l'ordre suivant : une vignette clinique (cfr explications données plus loin sur la rédaction d'une vignette clinique), une question et 5 propositions de réponses.

Exemples (référence n°1) :

Un adolescent de 15 ans a souffert de deux épisodes de choc anaphylactique sévère suite à des piqûres d'abeilles. Laquelle des interventions suivantes serait la plus adéquate ?

A. *Administration de corticostéroïdes durant l'été*
B. *Désensibilisation avec de l'extrait de venin d'abeille*
C. *Interdiction au patient de sortir de la maison durant les mois d'été*
D. *Tenue vestimentaire protectrice*
E. *Thérapie prophylactique de longue durée avec des antihistaminiques*

Un homme de 52 ans présente une augmentation de dyspnée et toux avec glaires purulentes depuis 2 jours. Il fume un paquet de cigarettes par jour depuis 30 ans. Sa température est de 37,2°C. Les bruits respiratoires sont assourdis avec quelques ronchis et sibilances. Le comptage des leucocytes est de 9000/mm3 avec une répartition normale. La coloration de Gram des glaires montre de nombreux neutrophiles et des diplocoques Gram négatif. Les radiographies du thorax montrent une hyper-inflation. Laquelle des propositions suivantes serait le diagnostic le plus probable ?

A. *Asthme*
B. *Bronchectasies*
C. *Bronchite*
D. *Embolie pulmonaire*
E. *Pneumonie à streptocoques*

Dans le **type R** (appelé aussi **QCM à appariement étendu**), l'énoncé se compose de 3 ou 4 parties. Dans le cas où il n'y a qu'une seule vignette pour la liste de propositions de réponse, la structure est identique aux questions de type A, la seule différence réside dans le fait qu'il y a plus de 5 propositions de réponse. Au cas où il y a plus d'une vignette pour une même liste de propositions de réponse, 4 parties composent alors l'énoncé : un thème, une liste de propositions, une instruction et plusieurs items ou vignettes.

Exemples :

Thème : Problèmes respiratoires chez l'enfant

Options :

A. *Asthme*
B. *Bronchiolite*
C. *Bronchite aiguë infectieuse*
D. *Coqueluche*
E. *Epiglottite*
F. *Laryngite striduleuse*
G. *Pneumonie*

Instruction :
Choisissez le diagnostic le plus probable pour les jeunes patients ci-dessous.

Vignettes cliniques

1. *Paul, 5 mois, est amené à la consultation par sa maman. Il tousse beaucoup depuis ce matin et sa maman a même entendu une respiration sifflante à certains moments. Son état général est altéré, il a de la température à 38,8°C et il n'a pas voulu manger beaucoup. C'est la première fois qu'il présente ces symptômes. A l'examen clinique, il est tachypnéique et l'expiration est discrètement sifflante.*
2. *Lise, 6 ans, tousse depuis deux jours. Depuis ce matin, elle présente une température élevée (39,5°C) et son état général est altéré. Elle a à peine voulu manger. A l'auscultation, vous entendez des crépitants au niveau de la base pulmonaire gauche.*

Dans **les 2 types de QCM (A et R)**, la majeure partie de l'information doit se trouver dans l'item. Celui-ci doit être assez étendu tandis que les options sont relativement brèves. L'item doit contenir toute l'information nécessaire pour répondre à la question. Aucune donnée additionnelle ne doit figurer dans les options.

De façon schématique, la structure doit se coller au modèle suivant (manuel de Case et Swanson, p 42, référence n°1).

Item avec une structure appropriée :

Item long

A.

B.

C. Propositions courtes

D.

E.

Item avec une mauvaise structure :

Item bref

A.

B.

C. Propositions longues

D.

E.

4.3 Recommandations pour la rédaction des QCM type A et type R

4.3.1 Le thème

En cas de **QCM à appariement étendu** avec plusieurs vignettes pour une liste de propositions, le thème annonce le sujet de la question. C'est par exemple, une plainte principale, une classe de médicaments, une pathologie, une démarche thérapeutique…

4.3.2 Le (les) item(s)

4.3.2.1 Structure de l'item

L'item est souvent une vignette clinique qui décrit un patient. Elle doit inclure un ou plusieurs des éléments suivants (dans le même ordre que celui-ci) :

- Age, sexe, profession, origine/ethnie
- Lieu du traitement
- Motif de la consultation
- Durée du symptôme
- Histoire clinique (=anamnèse) avec, éventuellement, les antécédents familiaux
- Statut, examen clinique
- Eventuellement, résultats des examens complémentaires
- Eventuellement, traitement initial, évolution…

L'usage de matériel de référence (table de vaccination, normes de résultats hématologiques, citations provenant de publications…) et d'illustrations (photos, radiographies, graphiques…) peut être approprié.

4.3.2.2 Différents thèmes possibles pour la rédaction d'items en rapport avec les tâches du médecin (référence n°1)

- La santé et le maintien de la santé

Ces items apprécient la capacité d'évaluer les facteurs de risque, de comprendre les données épidémiologiques, appliquer des mesures préventives… (tests de détection de maladies, vaccins et médecine du voyage, interventions d'urgence…).

- Mécanismes des maladies

Ces items évaluent la connaissance de la physiopathologie au sens large (étiologie, pathogenèse, histoire naturelle, histoire clinique, signes associés, complications, effets des interventions thérapeutiques…).

- Diagnostic

Les étudiants doivent interpréter l'histoire clinique, les signes physiques et les résultats d'examens de laboratoire, imagerie ou autres examens complémentaires dans le but de déterminer le diagnostic le plus probable ou l'étape suivante de la mise au point en vue d'établir un diagnostic.

- Traitement

Ces items évaluent les principes des traitements aigus et chroniques des patients ambulants et hospitalisés. Lors de la rédaction des questions sur le traitement, il faut se concentrer sur les aspects des soins qui sont importants pour le niveau de pratique de l'étudiant.

4.3.2.3 Règles à respecter lors de la rédaction des items

- Les items doivent être formulés de façon simple, claire et dans un langage accessible. Il faut éviter de les rendre artificiellement plus compliqués ou de les concevoir comme des pièges.
- Les items doivent contenir toute l'information nécessaire pour que le bon étudiant puisse répondre à la question sans regarder les propositions de réponses (c'est ce qui s'appelle la règle de « cacher les propositions »).
- La difficulté de l'item dépend non seulement de la complexité du problème clinique exposé et du niveau d'exigence cognitive nécessaire à sa résolution (compréhension, capacité de synthèse de la matière étudiée…) mais aussi de l'interprétation plus ou moins avancée du cas clinique exposé. L'étudiant devra montrer un niveau plus élevé de capacité d'interprétation, d'analyse et de synthèse à mesure que les données seront présentées sous une forme moins interprétée.
  Les 3 items suivants illustrent très bien cette notion (référence n°1).
  - Niveau d'exigence bas :
    Pas de vignette.
    Chez les enfants avec un syndrome néphrotique et une fonction rénale normale, quelle est l'anomalie rénale la plus probable parmi les propositions suivantes ?
    A. Glomérulonéphrite post streptococcique aiguë
    B. Purpura de Schönlein-Henoch avec néphrite
    C. Syndrome néphrotique avec changement minime
    D. Syndrome néphrotique due à une glomérulosclérose focale et segmentaire
    E. Syndrome hémolityque et urémique
  - Niveau d'exigence moyen :
    Vignette brève.
    Un enfant de 2 ans présente un œdème depuis une semaine. Sa tension artérielle est de 100/60 mmHg et présente un œdème généralisé et de l'ascite. Les dosages sériques sont les suivants : créatinine 0,4 mg/dL, albumine 1,4 g/dL et cholestérol 569 mg/dL. La tigette urinaire montre 4 + de protéines et ne présente pas de traces de sang. Quel est le diagnostic le plus probable parmi les propositions suivantes ?
    (même liste de propositions)
  - Niveau d'exigence élevé :
    Vignette étendue.
    Un enfant de 2 ans de race noire a présenté un gonflement au niveau des yeux et des chevilles la semaine passée. La tension artérielle est de 100/60 mmHg, le pouls est de 110/min et la fréquence respiratoire est de 28/min. En plus de l'œdème des yeux et des chevilles, il présente une distension abdominale avec onde liquidienne positive. Les dosages sériques sont les suivants : créatinine 0,4 mg/dL, albumine 1,4 g/dL et cholestérol 569 mg/dL. La tigette urinaire présente 4 + de protéines, pas de trace de sang. Quel est le diagnostic le plus probable parmi les propositions suivantes ?
    (même liste de propositions)

4.3.2.4 Règles supplémentaires à respecter lors de la rédaction des items pour des QCM à appariement étendu

- Les items doivent être de structure assez similaire. Il est préférable que le même type d'informations soit donné, par exemple, si un item contient des résultats de laboratoire, les autres items doivent en contenir aussi. Il vaut mieux aussi faire des vignettes avec des patients de plus ou moins même âge, même race et même genre de plaintes.
- Idéalement, lors de la rédaction des items, il faudrait écrire un item valable pour la majorité des propositions. C'est-à-dire que le nombre d'items par question doit être proche du nombre de propositions et que la plupart des propositions doivent faire partie des réponses correctes. Pour les propositions courantes ou importantes (par exemple : diagnostics courants ou diagnostics de première importance à ne pas rater), il peut y avoir plusieurs items.

4.3.3 L'instruction

L'instruction indique la relation entre les items et les propositions. Son rôle est donc de première importance.

4.3.3.1 Règles à respecter lors de la rédaction d'une instruction.

- L'instruction doit être bien formulée. Ceci permet qu'il n'y ait qu'une seule réponse juste à la question.
- La formulation doit être, de préférence, positive.
- L'instruction doit être formée d'une phrase entière. Eviter d'utiliser un début de phrase comme instruction et de proposer des fins de phrase comme propositions de réponse.
- Pour les **QCM à appariement étendu**, il n'y a qu'une seule instruction pour la série d'items du même thème.

4.3.3.2 Exemples d'instructions bien formulées.

- *Pour chacun des patients suivants, sélectionne le diagnostic le plus probable.*
- *Pour chacun des patients suivants, sélectionne l'étape de prise en charge suivante la plus appropriée.*
- *Pour chacun des patients suivants, sélectionne le symptôme additionnel qu'il pourrait présenter de façon la plus probable.*
- *Pour chacun des cas suivants, sélectionne l'intervention thérapeutique la mieux indiquée.*
- *Pour chacun des patients suivants, sélectionne l'étiologie la plus probable de leurs symptômes.*

4.3.4 La liste de propositions

4.3.4.1 Règles à respecter lors de la rédaction de la liste de propositions

- Les propositions de la liste doivent être aussi courtes que possible : un mot ou une phrase très brève.
- Les propositions doivent être homogènes (une liste de diagnostics, une liste de médicaments, une liste de localisations anatomiques…)
- Les propositions doivent avoir la même forme grammaticale.
- Eviter l'usage de verbes dans les propositions.
- Eviter les termes absolus comme toujours, jamais…
- Les propositions doivent être classées par ordre alphabétique. Exceptionnellement, elles peuvent être classées de façon thématique si cela semble plus logique.
- Chaque réponse fausse dans la liste est appelée « distracteur ». Les distracteurs doivent être de bons « concurrents » à la bonne réponse c'est-à-dire qu'ils ne doivent pas être de façon trop évidente incorrects. Il peut s'agir de faux raisonnements, de fausses idées largement répandues sur le sujet, de concepts dépassés… Plus la réponse est correcte et les distracteurs sont proches en termes de contenu et de justesse, plus l'item sera difficile.
- Les propositions peuvent être présentées sous forme de tableau (par exemple dans le cas de valeurs de laboratoire).
- Les propositions du type « toutes les réponses qui précèdent » ou « les réponse A et D sont toutes les deux correctes » ne doivent pas être utilisées. En effet, une réponse de ce type suppose qu'il y ait plusieurs bonnes réponses et enfreint la règle d'une seule bonne réponse ! Si, par contre, la proposition de ce type est fausse, elle peut être trouvée plus facilement car, il suffit qu'une des (2) réponses soient identifiées comme fausse.
- La proposition « aucune des réponses qui précèdent » devrait aussi être évitée.

4.3.4.2 Différence entre QCM de type A et QCM de type R

Dans le **type A**, la liste est limitée à cinq propositions.
Dans le **type R**, le nombre de propositions varie entre 6 et 26. Plus il y a de propositions, plus la difficulté du test est grande. Pour déterminer le nombre de propositions, il faudra tenir compte de l'objectif du test et du groupe qu'il cible.

4.3.4.3 Exemples de données pour la rédaction d'une liste de propositions

En fait, n'importe quelle donnée qui peut être énumérée dans une liste peut servir de base de propositions.
Exemples :

- En anatomie : artères, nerfs, muscles, structures anatomiques, composants du système locomoteur, structures cardiaques, composants du système nerveux central, segments de la moelle épinière…
- En physiopathologie : aminoacides, peptides, hormones, enzymes, molécules, protéines, lipides, vitamines, minéraux, structures endocrines, neurotransmetteurs, anomalies métaboliques, processus pathologiques, états physiopathologiques, anomalies électrolytiques, tests diagnostiques…
- En anatomopathologie : composants cellulaires, types de cellules, composants sanguins, types de tissus, …
- En génétique : caryotype, anomalies congénitales…

- En infectiologie : agents pathogènes, bactéries, champignons, virus, toxines, troubles immunitaires…
- En sémiologie : diagnostics
- En pharmacologie : médicaments, types de médicaments, alternatives de traitement…
- …

4.3.5 Défauts techniques à éviter lors de la rédaction de QCM (manuel de Case et Swanson, p 26, référence n°1)

Il existe deux types de défauts techniques les plus courants : ceux qui permettent de répondre correctement à la question sans connaître la réponse et ceux dus à des difficultés inutiles inclues dans les questions.
Des exemples de QCM mal rédigés seront donnés en illustration après l'explication des différents défauts.

4.3.5.1 Les défauts qui fournissent des indices (cues) pour identifier la bonne réponse (ou éliminer les distracteurs) et leurs « remèdes »

- Pistes grammaticales : pas de correspondance grammaticale entre un distracteur et l'énoncé. Ceci peut arriver facilement si l'énoncé est un début de phrase et que les propositions sont des fins de phrase (éviter absolument ce type d'énoncé). (Cfr exemples n°4 et n°8 ci-dessous.)
  → Toutes les propositions doivent correspondre à l'énoncé du point de vue grammatical. L'énoncé doit être une phrase entière et non un début de phrase à compléter.
- Pistes logiques : un sous-groupe de propositions est un ensemble exhaustif, c'est-à-dire que ce sous-groupe couvre à lui seul toutes les possibilités. (Cfr exemple n°7 ci-dessous.)
  → Eviter une liste exhaustive de propositions (qui formerait un ensemble à elles seules) à côté d'autres propositions.
- Usage de termes absolus : les termes « toujours » ou « jamais » figurent dans quelques propositions. (Cfr exemple n°5 ci-dessous.)
  → Eviter l'usage des termes absolus, il est rare qu'en médecine l'on puisse affirmer quelque chose de façon absolue avec un de ces termes. Il existe toujours des exceptions !
- La réponse est la proposition la plus longue, la plus spécifique ou la plus complète. (Cfr exemple n°8 ci-dessous.)
  → Les distracteurs devraient avoir une longueur et une différenciation aussi proches que possible de celles de la réponse correcte.
- Répétition de mots : répétition de mots ou de phrases dans l'énoncé et dans la réponse correcte. (Cfr exemple n°9 ci-dessous.)
  → Eviter les répétitions de mots ou de phrases dans l'énoncé et dans la bonne réponse. Ils mettront de suite l'étudiant sur la bonne piste.
- Stratégie de convergence : la réponse correcte contient la majorité des éléments qu'ont en commun d'autres propositions. (Cfr exemple n°10 ci-dessous.)
  → Eviter que la réponse correcte contienne le plus grand nombre d'éléments communs à d'autres réponses. Celle-ci aura plus de chance d'être correcte aux yeux de l'étudiant même médiocre.
- Pistes données dans d'autres items. Des informations sur la bonne réponse sont données dans d'autres items et/ou questions.

→ Eviter de donner des informations inutiles dans les items permet de réduire ce risque.

- Pistes liées à l'ordre des propositions. Si les propositions ne sont pas systématiquement classées par ordre alphabétique (ou par ordre thématique ou logique –ordre ascendant ou descendant par exemple-), le risque est grand de positionner préférentiellement la bonne réponse « au milieu » des propositions.
  → Classer les propositions par ordre alphabétique (ou thématique ou logique).

4.3.5.2 Les défauts liés à des difficultés inutiles inclues dans les questions

- Les propositions sont longues, compliquées ou doubles.
- Les données numériques sont exprimées de façon non uniforme. (Cfr exemple n°6 ci-dessous.)
- Les termes utilisés dans les propositions sont peu précis (ex : « parfois », « peu souvent », « normalement », « rarement », « presque jamais », « fréquemment », « souvent », « habituellement », « probablement »…). (Cfr exemple n°1 ci-dessous.)
- Les propositions sont classées de façon illogique. (Cfr exemple n°6 ci-dessous.)
- « Aucune des propositions citées » est utilisée comme proposition de réponse.
- Les énoncés sont trompeurs ou inutilement compliqués.
- La réponse à une question dépend de la réponse à une autre question en relation avec celle-ci.

4.3.5.3 Exemples de QCM mal rédigés

1. L'embolie pulmonaire :
   A. est toujours associée à de la fièvre
   B. est fréquemment associée à une toux productive
   C. n'est jamais observée chez les non-fumeurs
   D. peut toujours être confondues avec la pneumonie
   E. est traitée par l'administration d'héparine

Problèmes :
Mauvaise structure générale (pas de vignette clinique, propositions de réponse plus longues que le corps)
Pas de question introductive claire.
Usage de termes imprécis (fréquemment) et de termes absolus (toujours, jamais).
Non homogénéité des propositions (symptômes, diagnostics, facteurs, traitements)

2. Un homme de 55 ans a le souffle court et des expectorations purulentes. Il n'a pas d'antécédent d'hémoptysie ni de douleur thoracique. A plusieurs reprises au cours des derniers jours, il a eu des épisodes de sensation de froid ou de chaud mais il n'a pas éprouvé de frisson solennel. L'examen du thorax révèle une hyperinflation et des bruits respiratoires réduits sans matité ni râle crépitant, mais avec des sifflements dispersés. La radiographie thoracique est normale. La spirométrie révèle ce sui suit : VEMS : 1,68 (58% prévu), CVF : 2,12 (75% prévu).
   Dans la prise en charge du patient, vous suggéreriez :
   A. antibiotique intraveineux
   B. théophylline par voie orale
   C. thérapie bronchodilatatrice par inhalation
   D. cesser de fumer
   E. inhibiteurs de la leucotriène par voie orale

Problèmes :
Question d'introduction peu claire : La question vérifie-t-elle la première étape dans la prise en charge, les soins à long terme du patient ou la thérapie la plus efficace ?
Pas de classement alphabétique des propositions.

3. Dans le contexte d'un infarctus aigu du myocarde, parmi les médicaments suivants, lequel n'a pas été éprouvé comme réduisant la mortalité ?
   A. métoprolol par voie intraveineuse
   B. activateur du plasminogène des tissus-r par voie intraveineuse
   C. streptokinase par voie intraveineuse
   D. acide acétylsalicique
   E. nifédipine

Problèmes :
L'usage de la forme négative dans la question ne permet pas de savoir clairement ce que cette question chercher à savoir : la question vérifie-t-elle si le candidat sait que la nifédipine ne

réduit pas la mortalité ou que les autres médicaments le font effectivement ?
Pas de classement alphabétique des propositions.

4. Un homme de 60 ans, alcoolique, est amené au service des urgences par la police qui l'a trouvé inconscient sur la voie publique. Après s'être assuré que la voie respiratoire est libre, la première étape du traitement est l'administration intraveineuse de :
   A. examen du liquide céphalo-rachidien
   B. glucose avec vitamine B1 (thiamine)
   C. CT scan de la tête
   D. phénytoïne
   E. diazépam

Problèmes :
La question introductive est une phrase dont il manque la fin.
Pistes grammaticales : certaines propositions ne concordent pas grammaticalement avec la question.
Non homogénéité des propositions (examens complémentaires, médicaments)
Pas de classement alphabétique des propositions.

5. Chez les patients avec une démence avancée, type Alzheimer, le déficit de mémoire :
   A. peut être traité correctement avec la phosphatidylcholine (lécithine)
   B. pourrait être une séquelle d'une maladie de Parkinson précoce
   C. ne s'observe jamais chez les patients avec nœuds neurofibrillaires à l'autopsie
   D. n'est jamais sévère
   E. implique le système cholinergique

Problèmes :
Mauvaise structure générale (corps trop court, propositions longues).
Usage de termes absolus.
Non homogénéité des propositions.

6. Après un second épisode de salpingite, quelle est la probabilité qu'une femme soit stérile ?
   A. Moins de 20 %
   B. De 20 à 30 %
   C. Plus de 50 %
   D. 90 %
   E. 75 %

Problèmes :
Non homogénéité des propositions : soit que des intervalles, soit que des pourcentages précis.
Propositions classées de façon illogique (ordre croissant n'est pas respecté).

7. L'incidence du cancer du sein dans la population des femmes belges est en :
   A. augmentation
   B. diminution
   C. reste identique

D. est plus importante dans les milieux défavorisés
E. peut être diminuée avec une bonne campagne de prévention

Problèmes :
Piste logique : les propositions A, B et C couvrent toutes les possibilités (elles forment un ensemble exhaustif). Les étudiants qui débusquent facilement les indices pour identifier la bonne réponse savent que la bonne réponse se trouve dans les 3 premières propositions.
Mauvaise structure : corps court et certaines propositions longues.
Non homogénéité des propositions.

8. Le gain secondaire est :
   A. synonyme de simulation
   B. un problème fréquent dans les troubles obsessionnels compulsifs
   C. une complication d'une série de maladies et tend à prolonger la durée de beaucoup d'entre elles
   D. ne s'observe jamais dans le cas de lésion cérébrale organique

Problèmes :
Piste grammaticale : la proposition D n'est pas adaptée grammaticalement à l'énoncé.
L'énoncé n'est qu'un début de phrase.
La réponse correcte est la plus longue.
Non homogénéité des propositions.
Mauvaise structure générale : énoncé court, propositions longues.

9. Une patiente de 55 ans, fumeuse, suite à un retour de vacances (trajet en avion de plus de 12 heures) la veille, se plaint d'oppression thoracique, essoufflement, transpiration, toux avec hémoptysies, palpitations. A l'examen clinique, vous constatez que ses extrémités sont de couleur bleutée, son pouls est de 120 bpm, régulier, elle n'a pas de fièvre.
   Quel est le symptôme le plus discriminant pour établir votre diagnostic ?
   A. cyanose
   B. dyspnée
   C. hémoptysie
   D. sudation
   E. tachycardie

Problèmes :
Répétition de mots dans l'énoncé et dans les propositions.

10. Les anesthésiques locaux sont plus efficaces sous :
    A. la forme anionique et agissant depuis l'intérieur de la membrane nerveuse
    B. la forme cationique et agissant depuis l'intérieur de la membrane nerveuse
    C. la forme cationique et agissant depuis l'extérieur de la membrane nerveuse
    D. la forme neutre et agissant depuis l'intérieur de la membrane nerveuse
    E. la forme neutre et agissant depuis l'extérieur de la membrane nerveuse

Problèmes :
La réponse correcte contient la majorité des éléments qu'ont en commun d'autres propositions. (=stratégie de convergence)

5 Comment rédiger un examen de type QCM étape par étape ?

5.1 Définitions préalables

Il faut commencer par définir le domaine à tester.
Ensuite, il s'agit de définir l'objectif de l'évaluation.
Pour rappel, les différents objectifs d'une évaluation peuvent être :

- un objectif de sélection (détermination de notes de fin de cycle ou de fin de formation, décision de promotion)
- un objectif de formation (indiquer aux étudiants la matière importante et les matières où leur étude est insuffisante)
- un objectif de motivation (inciter les étudiants à étudier)
- un objectif de feed-back sur l'enseignement (indiquer aux enseignants le niveau de connaissances et compétences de leurs étudiants en vue d'évaluer la qualité de leurs cours)

5.2 Conception d'un plan directeur ou tableau de spécification

Il faut établir un « blueprint » ou « plan directeur » ou « tableau de spécification ».
Le Collège royal des médecins et chirurgiens du Canada définissent (référence n°4) celui-ci comme « des lignes directrices qui aident à planifier l'élaboration d'un examen. Les plans directeurs servent non seulement à déterminer le contenu et les compétences à évaluer, mais aussi à choisir la technique de mesure appropriée pour évaluer chaque compétence et à jauger la pondération des composantes, du contenu et des compétences dans l'examen. Ils favorisent la validité du contenu et procurent une compréhension générale du processus d'évaluation. Ils sont considérés confidentiels parce qu'ils énoncent spécifiquement le contenu qui sera présenté dans l'examen. »

5.3 Etablissement concret du contenu du plan directeur

Les évaluateurs doivent ensuite déterminer concrètement quel sera le contenu de l'examen.
Par exemple, si l'examen porte sur la capacité à diagnostiquer des problèmes cliniques communs, quelques évaluateurs doivent établir une liste des plaintes principales possibles de patient. Il convient de rédiger cette liste avant la réunion qui réunira tous les rédacteurs de l'examen.
Exemples de plaintes : masse abdominale, douleur abdominale, anémie, ascite, douleur de dos, douleur précordiale, confusion, toux, retard de développement, diarrhée, nausées, apparition facile d'hématomes, fatigue, fièvre, hémorragies digestives, céphalée, prurit, ictère, douleur articulaire, dysfonction sexuelle, dyspnée, lésions cutanées, syncope…

5.4 Formation d' un groupe d'enseignants à la rédaction d'items

Un groupe d'enseignants doit recevoir une formation à la rédaction d'items. Lors de cette formation, il faut donner une analyse brève de l'objectif de l'examen, expliquer la procédure à suivre pour la rédaction des items (règles à respecter) et donner des exemples qui illustrent bien les règles expliquées.

5.5 Rédaction des items, vérification, documentation

Il faut ensuite diviser le groupe d'enseignants en paires de rédacteurs d'items.
Pour continuer avec le même exemple déjà pris ci-haut, il faut attribuer à chaque paire 2 à 4 plaintes principales. Les étapes à suivre par les paires de rédacteurs sont par ordre chronologique :

- Identifier le thème général de leur question en fonction d'une plainte.
- Rédiger l'instruction introductive de leur question.
- Rédiger la liste de propositions (=une liste de diagnostics).
- Enfin, rédiger des vignettes cliniques de patients en rapport avec la liste de diagnostics en respectant la structure et l'ordre des éléments à inclure dans chaque item.

Chaque paire pourrait rédiger entre 20 et 40 énoncés… (Ecrire les énoncés sur ordinateur permet un grand gain de temps !)
Les rédacteurs doivent relire leurs questions en s'imaginant comment ils y répondraient eux-mêmes. Ils doivent vérifier que les règles de rédaction explicitées dans ce manuel sont respectées et qu'ils ont évité de fournir des « cues ».
Ils documentent leurs items en précisant l'objet de chaque vignette par quelques mots clés, en citant si possible une ou plusieurs références de la littérature (permettant de justifier la réponse correcte) et en mentionnant à quel objectif ou domaine (cfr le blueprint), l'item correspond.
(L'idéal est de laisser « reposer » les items rédigés pendant une semaine et les relire ensuite.)

5.6 Vérification en groupe des items

Après avoir fait réunir à nouveau les paires de rédacteurs, chaque auteur doit soumettre aux autres les items qu'il a écrits en les lisant, par exemple, à haute voix. Les autres essayeront de répondre. Il faut aussi soumettre au groupe la liste de propositions et la corriger si nécessaire.

5.7 Vérification par des tiers et des experts

Après avoir mis en page la liste des questions rédigées, il faut les soumettre à des tiers sans indiquer la réponse correcte. Il leur sera demandé de répondre aux questions comme s'ils passaient l'examen. Ceci permettra d'éliminer les questions pour lesquelles plusieurs propositions sont choisies comme réponse correcte.
Les questions doivent aussi être vérifiés par des experts de 2 types. D'abord, des experts du domaine concerné vérifient le contenu : justesse et appartenance du thème au programme de formation. Ensuite, des experts en examens QCM vérifieront les questions du point de vue de la forme et de la langue.

5.8 Acceptation ou refus des items

Les responsables de la réalisation des questions reçoivent les questions avec les remarques des experts et décident d'accepter, de refuser ou de demander de modifier les questions.

5.9 Elaboration de l'examen

Il s'agit de sélectionner un échantillon d'items pour chaque plainte (cfr l'exemple ci-haut) et de garder les autres pour des examens ultérieurs.
Ensuite, l'élaboration de l'examen suppose la rédaction d'une instruction standardisée décrivant la méthode la plus sûre et la plus économique de répondre aux questions (cfr exemple donné en annexe 2). Elle doit aussi informer les étudiants de la procédure d'évaluation des réponses. Il faut préciser si les mauvaises réponses seront sanctionnées ou s'il vaut mieux répondre au hasard en cas d'incertitude. Dans le cas des QCM de type A et R, il est recommandé de ne pas sanctionner les mauvaises réponses. L'étudiant aura donc toujours intérêt à choisir une proposition de réponse même s'il doute de sa réponse.
Il est conseillé de commencer l'examen par 1 ou 2 items faciles pour calmer l'angoisse de début d'examen des étudiants.

6 Interprétation des résultats obtenus en vue de l'analyse des items

Il est très important d'avoir un retour des résultats de l'examen par QCM. L'analyse des résultats pour chaque item permet d'évaluer la qualité des questions. Il faut donc envoyer un rapport avec les résultats de l'analyse des items aux professeurs après chaque examen à choix multiple.

L'approche classique consiste en l'analyse des indices de difficulté et de discrimination des items. Cette approche est basée sur le prémisse que chaque item mesure la même dimension, le même trait (par exemple le raisonnement clinique) ce qui est discutable dans les études (para)-médicales.

Nous proposons une analyse des items faite sur base de l'étude des profils de réponses.

Pour illustrer cette notion, nous reprenons un exemple détaillé dans le manuel de Case et Swanson (référence n°1, p 107).

Les étudiants qui ont passé l'examen ont été divisés en un groupe « supérieur » et un groupe « inférieur » selon leur résultat général à tout l'examen. Si le groupe contient un nombre réduit d'étudiants, ces 2 groupes sont formés respectivement par les 50 % d'étudiants ayant le mieux réussi et les 50 % des étudiants ayant le moins bien réussi. Si, par contre, le nombre d'étudiants qui ont passé l'examen est élevé, le groupe « supérieur » peut alors inclure seulement les 25 % des étudiants ayant les meilleurs résultats et le groupe « inférieur » les 25 % des étudiants ayant les moins bons résultats.

Le rapport d'analyse des items va montrer pour chaque option de réponse, le pourcentage des étudiants de chaque groupe qui ont choisi cette option.

Exemples :

Pour chaque exemple, le pourcentage d'étudiants qui ont choisi chaque proposition est donné. La ligne « Total » reprend tous les étudiants qui ont passé l'examen, la ligne « Supérieur » reprend le groupe « supérieur » (cfr explication donnée ci-haut) et la ligne « Inférieur », le groupe « inférieur ». L'astérisque à côté de la lettre symbolisant l'option montre que c'est l'option supposée être la bonne réponse.

Par exemple, pour l'item n°1, 1 % du groupe supérieur a choisi l'option A, 1 % l'option B, 91 % l'option C, 4 % l'option D, 1 % l'option E et 2 % l'option F. Par contre dans le groupe inférieur : 20 % a choisi l'option A, 6 % l'option B… La réponse supposée correcte est la réponse B.

Item n°1

		Options					
		A	B*	C	D	E	F
Groupes	Supérieur	1	**1**	91	4	1	2
	Inférieur	20	**6**	51	14	6	3
	Total	**9**	**2**	**76**	**8**	**3**	**2**

Interprétation :
Ceci est l'exemple typique d'une erreur au niveau de la détermination de la réponse correcte. Si la réponse correcte est l'option B, seulement 2 % des étudiants ont répondu correctement. La réponse correcte semble être la C. Un expert devra réviser le contenu de la question pour en être sûr. S'il confirme que la réponse correcte est la C, alors l'item peut être considéré comme une bonne question, à condition d'enregistrer l'option C comme étant la réponse correcte.

Item n°2

		Options					
		A	B	C*	D	E	F
Groupes	Supérieur	0	1	**90**	3	3	3
	Inférieur	0	1	**60**	25	8	6
	Total	**0**	**1**	**74**	**12**	**7**	**6**

Interprétation :
90 % du groupe supérieur et 60 % du groupe inférieur ont sélectionné la réponse correcte (la C). Ceci est un bon résultat de profil d'item. Cependant, il pourrait y avoir une révision des options A et B avant de réutiliser cet item dans un prochain examen parce que très peu d'étudiants ont choisi ces options de réponse. L'idéal serait de trouver des distracteurs plus « plausibles ».

Item n°3

		Options					
		A	B	C*	D	E	F
Groupes	Supérieur	44	1	**50**	2	1	2
	Inférieur	20	15	**21**	22	20	2
	Total	**32**	**7**	**34**	**14**	**11**	**2**

Interprétation :
50 % du groupe supérieur et 21 % du groupe inférieur ont choisi la réponse supposée être correcte (la C). Cet item est très difficile et est probablement mal rédigé. Beaucoup d'étudiants du groupe supérieur ont choisi l'option A. Soit l'item est mal rédigé, soit l'option A est potentiellement une bonne réponse… Cet item doit être revu par les experts.

Item n°4

		Options					
		A	B	C*	D	E	F
Groupes	Supérieur	18	10	**51**	17	2	2
	Inférieur	24	24	**21**	25	4	2
	Total	**22**	**17**	**34**	**22**	**3**	**2**

Interprétation :
La répartition des groupes supérieur et inférieur pour l'option C est quasi identique à l'item n°3. Mais cet item est peut-être bien rédigé. La différence entre les 2 items est que, pour celui-ci, les étudiants ne connaissent pas la bonne réponse. C'est-à-dire que la répartition se fait entre les différents distracteurs. Evidemment, il est préférable de réviser les options A, B et D afin de contrôler qu'elles sont bien des distracteurs et non des réponses potentiellement correctes et qu'elles sont bien rédigées.

7 Comment établir une note de passage ?

7.1 Définition et considérations générales

Une note de passage est le score obtenu au test qui séparera les étudiants ayant réussi le test de ceux qui ont échoué. La note de passage est établie suite à une décision pédagogique. Ceci demande un critère de jugement. Ce critère de jugement est arbitraire mais il faut éviter qu'il soit choisi de façon fantaisiste. A moins qu'il n'existe une raison spécifique de limiter le nombre d'étudiants qui réussissent, une note de passage basée sur la maîtrise que l'étudiant possède de la matière est préférable à une note de passage basée sur un indice d'échec. Il est conseillé de faire participer au processus d'établissement d'une note de passage plusieurs évaluateurs bien informés de la procédure. Les évaluateurs devront entrer en possession des données de performance des étudiants à un moment donné de l'établissement des notes de passage.
Il existe différentes manières d'établir une note de passage. Elle peut être relative, absolue ou être un compromis entre ces deux méthodes.

7.2 Méthode relative

Une note de passage relative se base sur la performance du groupe qui passe l'examen. Les étudiants réussissent ou échouent en fonction du niveau du groupe auquel ils appartiennent. Si le niveau de performance du groupe est élevé, c'est à dire que la moyenne des résultats est élevée, la note de passage sera plus élevée que si le niveau de performance du groupe est bas (moyenne des résultats plus basse).
Par exemple, il peut être décidé que les étudiants qui ont obtenu un résultat inférieur à 2 déviations standards de la moyenne des résultats ont échoué.
Il est déconseillé d'utiliser une méthode relative. En effet, elle désavantage fortement les étudiants appartenant à des groupes (années de promotions par exemple) performants. Un étudiant moyen qui a les compétences requises pour être un bon professionnel de la santé risquerait d'échouer parce qu'il fait partie d'un groupe d'étudiants particulièrement brillants.

7.3 Méthode absolue

Une note de passage absolue se base sur le niveau de performance de l'étudiant sans tenir compte de celui des autres étudiants. Il n'y a pas de comparaison entre le niveau de performance de l'étudiant et celui de son groupe lors de la décision de réussite ou d'échec. Il peut, par exemple, être décidé que l'étudiant doit obtenir une bonne réponse à 60 % des questions pour réussir l'examen.

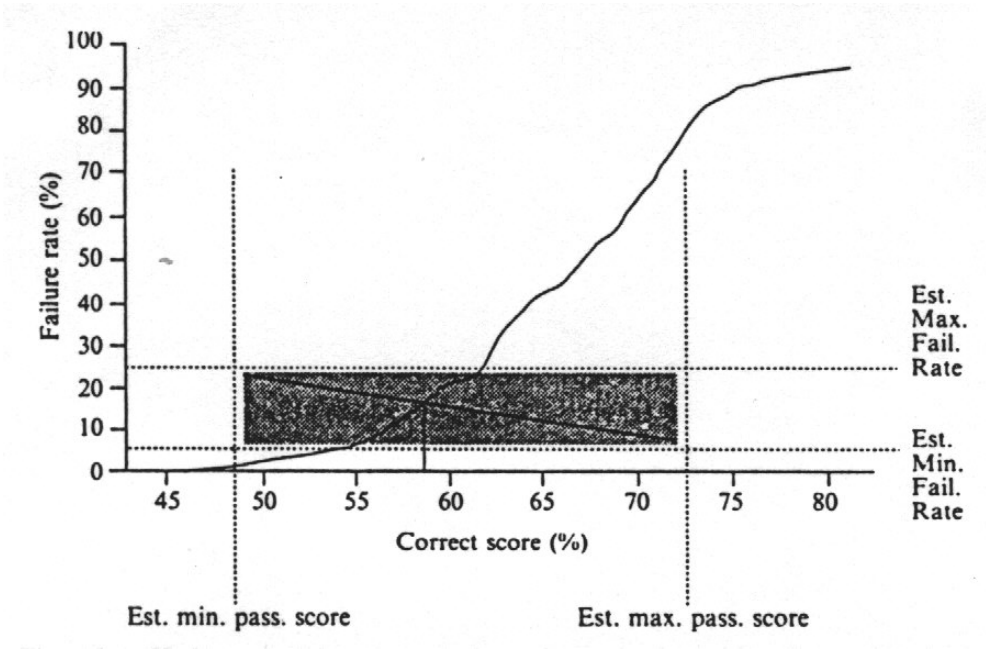
Une méthode couramment utilisée est celle d'**Angoff**. Les membres du jury vont évaluer pour chaque question la probabilité qu'un étudiant « borderline » (un étudiant qui serait à la limite entre réussir et échouer) réponde correctement. La probabilité va de 0 à 1. Après avoir révisé chaque question et estimé cette probabilité, la note de passage est calculée en fonction de la somme des probabilités obtenue.

7.4 Compromis entre une méthode absolue et une méthode relative

La méthode d'Hofstee est un compromis entre une méthode absolue et une méthode relative. Après avoir révisé une copie de l'examen, les évaluateurs vont définir les valeurs des notes de passage acceptables suivantes :

- Pourcentage minimum acceptable d'étudiants refusés (indice minimum d'échec)
- Pourcentage maximum acceptable d'étudiants refusés (indice maximum d'échec)
- Score le plus bas qui permettrait à un étudiant de réussir l'examen (point minimal de réussite)
- Score le plus élevé nécessaire pour qu'un élève réussisse (point maximal de réussite)

Un graphique avec le pourcentage d'échecs en ordonnée et le pourcentage de réponses correctes en abscisse est tracé. Les quatre points cités ci-haut sont reliés pour former un rectangle.
Une courbe représentant le pourcentage de réussites en fonction d'une note de passage définie arbitrairement est dessinée. (Si on fixe la note de passage suivante : il faut 80 % de réponses correctes pour réussir, il y aura 90 % d'étudiants en échec. Si la note de passage exigée est de 60 % de réponses correctes, il y aura 20 % d'étudiants en échec…)
L'intersection entre la diagonale partant du coin supérieur gauche du rectangle vers le coin inférieur droit et la courbe détermine le standard.



Graphe repris de la thèse de J. Degryse (référence n°5)

8 Conclusion

L'évaluation par QCM a largement prouvé son efficacité dans le domaine des sciences médicales. Le respect de règles bien établies permet une évaluation valide et fiable. Néanmoins, cela demande un travail de préparation et de contrôle a posteriori important. Pour de plus amples détails, nous vous encourageons à consulter les références sur lesquelles nous nous sommes basées. La majorité de ces références sont disponibles sur internet. Les adresses pour les télécharger sont indiquées dans la bibliographie.

9 Bibliographie

1. Case SM, Swanson DB. Constructing Written Test Questions For the Basic and Clinical Sciences. Third Edition. Philadelphia : National Board of Medical Examiners, 2002.
   http://www.nbme.org/publications/index.html

2. Krebs R. Mode d'emploi pour la création de questions et d'examens à choix multiples pour l'enseignement de la médecine. Berne : Institut für Medizinische Lehre, Universitat Bern, 2004.
   http://www.iawf.unibe.ch/aae/documents/methodik/Manuel_QCM.pdf

3. Wood T, Cole G. L'élaboration de questions à choix multiples pour les examens menant au certificat du Collège royal des médecins et chirurgiens du Canada. Le Collège royal des médecins et chirurgiens du Canada, 2001.
   http://www.medbev.umontreal.ca/docimo/cotonou.htm

4. Cole G, Lee C, Mikhael N. L'élaboration d'un plan directeur d'examen ProMEDS du Collège royal. Le Collège royal des médecins et chirurgiens du Canada, 2003.
   http://www.medbev.umontreal.ca/docimo/cotonou.htm

5. Degryse J. De eindproef huisartsgeneeskunde. Ontwikkeling en evaluatie van de certificatieprocedure voor huisartsen in Vlaanderen. Leuven : Katholieke Universiteit Leuven. 2003.

10 Annexes

10.1 Formulaire pour la rédaction d'items (manuel de Case et Swanson, p 93, référence n°1)

Thème général : _____
(par exemple : une plainte que présente un patient)

Instruction introductive : _____
(par exemple : pour chaque patient qui présente de la fièvre, sélectionne le diagnostic le plus probable)

Propositions :

A	N
B	O
C	P
D	Q
E	R
F	S
G	T
H	U
I	V
J	W
K	X
L	Y
M	Z

Items : à écrire sur une feuille séparée.

10.2 Aide-mémoire utile

AIDE-MEMOIRE UTILE A L'ELABORATION DE QCM TYPE A+ ET TYPE R		
	QCM type A+	**QCM type R**
structure générale	1) un item ou vignette clinique	1) un thème
	2) une question	2) une liste de propositions (6 à 26)
	3) 5 propositions de réponse	3) une instruction
		4) plusieurs items ou vignette clinique
	item ou vignette clinique : structure longue	
	propositions : structure courte	
thème		annonce le sujet
item (ou vign. clin.)	description d'un patient	
	formulation claire, simple	
	information complète (cfr règle "cacher les propositions")	
	difficulté = complexité problème clinique, interprétation + ou - avancée	
		structures similaires
		nombre items = + ou - nombre propositions
instruction		indique la relation entre items et propositions
		formulation claire
		de préférence positive
		une phrase entière
propositions	aussi courtes que possible	
	homogènes	
	même forme grammaticale	
	éviter l'usage de verbes	
	éviter les termes absolus (toujours, jamais…)	
	classement par ordre alphabétique (ou thématique si plus logique)	
	bons distracteurs (= bons concurrents à la bonne réponse)	
	une seule réponse correcte	
éviter les défauts techniques		
indices ("cues")	pistes grammaticales	
	pistes logiques	
	usage de termes absolus	
	la proposition correcte est la plus longue, la plus complète	
	répétition de mots ou de phrases dans l'énoncé et dans la bonne réponse	
	la réponse correcte contient plusieurs éléments cités dans les autres propositions	
	pistes données dans d'autres items	
	pistes liées à l'ordre des propositions	
difficultés inutiles	propositions longues, compliquées, doubles	
	données numériques non uniformes	
	termes peu précis (parfois, rarement, peu souvent, habituellement…)	
	propositions classées de façon illogique	
	énoncés trompeurs, inutilement compliqués	
	la réponse à une question dépend de la réponse à une autre question	
à éviter dans les propositions	"toutes les réponses qui précèdent"	
	"les réponses A et D sont toutes les 2 correctes"	
	"aucune des réponses qui précèdent"	

10.3 Exemple d'instruction standardisée pour un QCM de type R

INSTRUCTIONS

Ce test est composé de 50 questions de type QCM à appariement étendu et comprend 160 vignettes cliniques au total.

Lisez attentivement l'instruction qui accompagne chaque question.

Il n'y a qu'une seule réponse correcte par vignette clinique.

Une option de réponse peut correspondre à une seule vignette clinique, à plusieurs vignettes cliniques ou à aucune.

Vous ne serez pas pénalisé pour les réponses fausses : il vaut donc mieux tenter votre chance même si vous n'êtes pas sûr de votre réponse.

Travaillez à votre aise mais ne vous arrêtez pas trop longtemps sur les questions qui vous posent problème : vous pourrez toujours y revenir par la suite.

Noircissez bien la case correspondant à votre réponse au crayon sur les feuilles ad hoc.

Vous disposez de 3 heures.

Bon travail !

10.4 Exemples de QCM bien rédigés

Certains exemples ont été tirés des références citées dans la bibliographie.

10.4.1 QCM à appariement étendu :

Thème : Syndromes thoraciques

Options :

A. Angor
B. Bronchite aiguë
C. Crise d'asthme
D. Crise de panique
E. Embolie pulmonaire
F. Infarctus du myocarde
G. Œdème pulmonaire aigu
H. Oesophagite
I. Péricardite aiguë
J. Pleurésie
K. Pneumonie
L. Pneumothorax
M. Rupture d'un anévrisme de l'aorte
N. Syndrome de Tietze

Instruction :
Choisissez pour chaque patient ci-dessous l'hypothèse diagnostique la plus probable.

Vignettes cliniques :

1. Vous êtes appelé en urgence auprès d'un homme de 45 ans. Il présente des difficultés respiratoires depuis quelques heures. Il est assis au bord de son lit. Sa fenêtre est grand ouverte. Il fait de grands efforts pour respirer. Il est très inquiet et s'appuie sur ses deux bras de manière crispée. Ce n'est qu'avec difficulté qu'il parvient à aligner quelques mots. Il n'a pas de fièvre mais son pouls est rapide. A l'auscultation, vous entendez des sibilances (wheezing) au niveau des deux champs pulmonaires.

2. Une dame obèse dans la quarantaine, vient vous consulter. Elle est en processus de deuil car son mari est décédé brutalement il y a quelques mois. Dernièrement, vous lui aviez prescrit des bas de contention à cause de signes d'insuffisances veineuse au niveau des jambes. Lorsqu'elle entre dans le cabinet, elle se laisse tomber dans le fauteuil pour retrouver sa respiration. Elle est dyspnéique au moindre effort. Elle respire rapidement et superficiellement. Son pouls est rapide, avec de nombreuses extrasystoles. Elle pense que cela va passer en se reposant. Elle ressent une légère douleur dans la poitrine. Elle toussote précipitamment en permanence, ce qui ramène parfois une petite expectoration rose pâle.

3. Une femme vous demande d'examiner en urgence son mari de 28 ans, pour une grippe sévère. Elle s'inquiète de son état. L'homme est alité, allongé sur le côté droit. Il a de la fièvre à 40,5°C, frissonne et transpire abondamment. Il dit avoir mal dans tous les muscles, mais surtout au niveau de la poitrine. La douleur augmente à chaque inspiration et lors des mouvements. Vous auscultez le thorax et vous remarquez de fines crépitations intermittentes au niveau précordial, rythmées par le pouls.

4. Une femme de 45 ans est traitée de manière symptomatique pour un syndrome grippal. Après quatre jours, vous l'examinez à nouveau. Peu de changements sont apparus au niveau de la symptomatologie. Après discussion avec la patiente, le traitement est poursuivi sans modification. Deux jours plus tard, elle vous rappelle en urgence. Elle est assise, moite et transpirante, dans le fauteuil. Elle est fort affaiblie. La fièvre reste élevée et elle tousse toujours. La fréquence respiratoire est accélérée. A l'auscultation, vous entendez des râles crépitants au niveau du poumon gauche, surtout localisés au niveau de la base.

Thème : infectiologie

Options :

A. Adénovirus
B. Aspergillus fumigatus
C. Bacillus anthracis
D. Candida albicans
E. Chlamydia psittaci
F. Coccidioides immitis
G. Coronavirus
H. Coryneacterium diphteriae
I. Coxiella burnetii
J. Coxsackievirus
K. Epstin-Barr Virus
L. Haemophilus influenzae
M. Histoplasma capsulatum
N. Mycobacterium tuberculosis
O. Mycoplasma pneumoniae
P. Neisseria gonorrhoeae
Q. Neisseria meningitidis
R. Pneumocystitis carinii
S. Rhinovirus
T. Streptococcus pneumoniae
U. Streptococcus pyogenes (groupe A)

Instruction :
Veuillez indiquer pour chacun des patients suivants présentant un état fébrile le microorganisme qui a le plus de chances d'être l'agent responsable de la maladie décrite.

Vignettes cliniques :

1. Un enfant de 7 ans présente un état hautement fébrile et des douleurs pharyngées. Le pharynx est hyperémié, l'amygdale droite tuméfiée est recouverte d'un exsudat mucoïde blanchâtre. Il y a une adénopathie sous-angulomaxillaire droite douloureuse. La culture du frottis sur plaque d'agar montre de nombreuses petites colonies β-hémolytiques inhibées par la bacitracine.
2. Un patient de 18 ans souffre depuis une semaine d'un état fébrile avec des douleur pharyngées et un malaise général. Les amygdales sont tuméfiées et recouvertes d'un exsudat. On observe aussi la présence d'adénopathies cervicales diffuses et d'une splénomégalie. On trouve une lymphocytose avec des lymphocytes atypiques. Le test pour les anticorps hétérophiles est positif.

10.4.2 QCM de type A+ :

1. Une femme de 60 ans éprouve des difficultés à sortir de son fauteuil et à se tenir droite. Elle n'a en revanche aucune peine à marcher sur sol plat et parvient à fléchir ses hanches sans aucune difficulté.

   Lequel des muscles suivants est le plus probablement déficient ?

   A. M. grand fessier
   B. M. grand psoas
   C. M. moyen fessier
   D. M. obturateur externe
   E. M. obturateur interne

2. Un homme de 65 ans, obèse, souffre de troubles respiratoires progressifs. La spirométrie donne les résultats suivants :
   - diminution des volumes pulmonaires
   - rapport normal entre volume maximal expiré à la 1ère seconde et capacité vitale (FEV1/VK)
   - réduction sévère de la compliance pulmonaire
   La gazométrie révèle une hypoxémie artérielle et une hypocapnie au travail.

   Quelle est la cause la plus vraisemblable de ce tableau clinique ?

   A. une augmentation du débit circulatoire pulmonaire en raison d'un shunt artérioveineux
   B. une bronchite chronique
   C. un emphysème pulmonaire obstructif
   D. une fibrose interstitielle diffuse des poumons
   E. une obésité

3. Le menu d'un banquet était composé d'un rôti de poulet, de pommes de terre rôties, de petits pois, d'un éclair au chocolat et d'un café. La majorité des convives présentent des nausées deux heures après le repas, se mettent à vomir et à se plaindre de douleurs de abdominales.

   Lequel de ces micro-organismes suivants a-t-on le plus de chances de trouver dans les analyses des aliments ?

   A. Escherichia coli
   B. Proteus mirabilis
   C. Salmonella typhimurium
   D. Staphylococcus aureus
   E. Streptococcus faecalis

4. Lors d'un check-up de routine, une femme de 75 ans qui a une activité modérée présente une concentration de cholestérol sérique total de 208 mg/dl et une concentration de cholestérol HDL sérique de 70 mg/dl. L'ECG ne montre aucune anomalie.

   Laquelle des recommandations suivantes serait la plus adéquate ?

   A. Augmentation de l'ingestion de fibres
   B. Diminution de l'ingestion de cholestérol
   C. Diminution de l'ingestion d'hydrates de carbone simples
   D. Diminution de l'ingestion de graisses saturées
   E. Ne pas faire de changement dans le régime.

5. Un adolescent de 15 ans, sans antécédent médical, présente une douleur périombilicale spasmodique. Après quelques heures, la douleur migre vers le quadrant inférieur droit et devient constante. Il vomit plusieurs fois. Il est amené au service d'urgences. L'abdomen est sensible à la palpation profonde du quadrant inférieur droit. Les radiographies de thorax et abdomen sont normales. Le nombre de leucocytes dans l'analyse sanguine est de 15 000/mm3. L'analyse d'urine montre 3 leucocytes/champ.

   Laquelle des propositions suivantes serait la prise en charge la plus adéquate ?

   A. CT Scan de l'abdomen
   B. Cystographie et pyélographie intraveineuse
   C. Exploration chirurgicale de l'abdomen
   D. Lavement baryté
   E. Traitement de soutien à domicile avec retour immédiat aux urgences si la douleur augmente

6. Une infirmière est hospitalisée pour une appendicectomie dans la clinique où elle travaille. Une semaine après qu'elle soit sortie de clinique, l'assistant de l'administrateur de l'hôpital demande au chirurgien quel était le diagnostic définitif.

   Laquelle des propositions suivantes serait la réponse la plus adéquate de la part du chirurgien ?

   A. Refuser de répondre parce que l'administrateur n'est pas médecin
   B. Refuser de répondre parce que l'information est confidentielle
   C. Répondre en raison du risque de la possibilité de diffusion de fausse information au sujet de la patiente.
   D. Répondre parce que cela permettra une accélération du management de l'assurance de la clinique.
   E. Répondre parce que, en tant qu'employé de la clinique, l'administrateur a accès à l'information sur les patients.